

## USING PAGE-VIEW DATA TO PROJECT DEMAND FOR AN ITEM

### TECHNICAL FIELD

[0001] The present invention is directed to the fields of electronic commerce and statistical analysis.

### BACKGROUND

[0002] Every merchant is in the business of making items available for purchase by purchasers. Many merchants do or would find it helpful to have an accurate forecast of future purchasing activity for some or all of these items. Such a forecast indicates what quantities of an item will be sold at each of a number of future times, at each of one or more merchant locations, such as stores or distribution centers.

[0003] Such a forecast, if accurate, can help ensure that adequate resources are available to facilitate the forecasted purchases, such as inventory in the item, storage capacity for the inventory, inventory in complementary items (e.g., batteries of the type used in the item), or workers needed to support purchasing activities. Such a forecast may also enable the merchant to make a more accurate projection of future financial performance, allowing the merchant to better plan for various cash flow issues. Such a forecast may also enable merchants to better target promotional initiatives, such as advertising, item placement, sales, etc.

[0004] Unfortunately, it is not often possible to produce accurate projections, as conventional approaches to constructing them have significant disadvantages. In a first conventional approach, future purchasing activity is projected based upon past purchasing activity of the same item. Such projections are unfortunately

merely rough guesses, as there is sometimes no meaningful relationship between past purchasing activity and future purchasing activity. Accordingly, this approach can require substantial manual "sales case" analysis to achieve anything approaching a significant likelihood of accuracy. Also, it is impossible to make projections using the first approach where past purchasing activity does not exist—as is true for new items that have not yet been available to purchase—or where past purchasing activity is limited by some factor external to demand, such as lack of inventory during a stock-out period, or inability of customers to purchase when a credit card authorization system is out of service.

- [0005] In a second conventional approach, future purchasing activity is projected based upon past purchasing activity of an analogous item. When pursuing this approach it can be difficult or impossible to identify analogs whose purchasing activity the item in question will mirror. Further, even in cases in which it is possible to identify a useful analog, it is often difficult to match the analog's purchasing activity against that of the item in question.
- [0006] In view of the substantial shortcomings of conventional approaches to projecting future purchasing activity for an item, a more effective approach to doing so would have significant utility.

#### BRIEF DESCRIPTION OF THE DRAWINGS

- [0007] Figure 1 is a high-level block diagram showing a typical environment in which the facility operates.
- [0008] Figure 2 is a data flow diagram depicting a first approach used by the facility to generate a blended purchasing forecast for an item.
- [0009] Figure 3 is a graph showing the facility's generation of a sample blended purchasing forecast in accordance with the first approach.
- [0010] Figure 4 is a data flow diagram depicting a second approach used by the facility to generate a blended purchasing forecast for an item.
- [0011] Figure 5 is a graph showing the facility's generation of a sample blended purchasing forecast in accordance with the second approach.

[0012] Figure 6 is a data flow diagram showing a third approach used by the facility to generate a purchasing forecast for an item.

## DETAILED DESCRIPTION

[0013] A software facility for projecting future purchasing activity of an item based on past browsing activity for the item ("the facility") is described.

[0014] In some embodiments, for a web-based merchant offering items for sale, for each item, the facility identifies types of HTTP requests that, when received from client computer systems being used by customers, constitute a browsing activity for the item. The identified requests can include such requests as the following: a request for a page containing information about multiple items that include that item (such as an item category page), a request for a page containing information only about that item (such as an item detail page), a request for a page containing further information about that item (such as a page containing item reviews), a request for signing up to be notified when the item becomes available, a request adding the item to a shopping cart or a gift registry, a search request specifying a query string for matching items of interest, etc. The facility typically extracts and counts requests of these types from web server logs produced by the merchant's web site to generate past browsing activity metrics for each item.

[0015] In some embodiments, the facility uses the past browsing activity metrics to project future browsing activity, then converts the projection of future browsing activity into a projection of future purchasing activity. Alternatively, in some embodiments, the facility uses the past browsing activity metrics to directly project future purchasing activity. In some embodiments, the facility blends the projection of future purchasing activity produced directly or indirectly from the past browsing activity metrics with a parallel projection of future activity generated from past purchasing activity. In some embodiments, the facility projects future purchasing activity directly from past browsing activity and past purchasing activity.

[0016] In many cases, the facility produces forecasts of future purchasing activity that are significantly more accurate than conventional approaches. The facility can generally produce a future purchasing activity forecast for any item for which browsing activity information is available, irrespective of the availability or cohesiveness of past purchasing activity information, and irrespective of whether analogous items exist or have been identified.

[0017] Figure 1 is a high-level block diagram showing a typical environment in which the facility operates. The block diagram shows several client computer systems, such as client computer systems 110, 120, and 130. Each of the client computer systems has a web client computer program for browsing the World Wide Web, such as web clients 111, 121, and 131. The client computer systems are connected via the Internet 140 to a web merchant server computer system 150 hosting the facility. Those skilled in the art will recognize that client computer systems could be connected to the server computer system by networks other than the Internet, however. By using the World Wide Web to communicate with the web merchant server computer system, users of the client computer systems can perform browsing and purchasing activities with respect to items offered for sale by the web merchant.

[0018] The web merchant server computer system 150 contains a memory 160. The memory 160 preferably contains the facility 161 for projecting future purchasing activity of an item. The facility typically constructs a future purchasing activity forecast 162 using information about past browsing activity 163, or information about past browsing activity in conjunction with information about past purchasing activity 164. Information about these activities is typically extracted from a web log 165 produced by a web server computer program 166 for delivering web pages in response to requests from web clients. Those skilled in the art will appreciate that information about past browsing activity and/or past purchasing activity may be derived from a variety of different sources, in a variety of different manners. While items 161-166 are preferably stored in memory while being used, those skilled in the art will appreciate that these items, or portions of

them, maybe be transferred between memory and a persistent storage device 172 for purposes of memory management and maintaining data integrity. The server computer system further contains one or more central processing units (CPU) 171 for executing programs, such as programs 161 and 166, and a computer-readable medium drive 173 for reading or writing information or installing programs such as the facility from computer-readable media, such as a floppy disk, a CD-ROM, or a DVD.

[0019] While various embodiments are described in terms in the environment described above, those skilled in the art will appreciate that the facility may be implemented in a variety of other environments including a single, monolithic computer system, a distributed system, as well as various other combinations of computer systems or similar devices connected in various ways.

[0020] Figure 2 is a data flow diagram depicting a first approach used by the facility to generate a blended purchasing forecast for an item. The facility begins with two types of information about past activity with respect to the item: browsing history information 201 and purchasing history information 211 documenting browsing and purchasing activities performed at a web merchant.

[0021] The browsing history information 201 describes browsing activities that have recently been performed by customers with respect to the item. Examples of these activities are discussed above. In some embodiments, the browsing history information constitutes an array indexed in one dimension by time. As discussed further below, in some embodiments, the array is indexed in another dimension by activity type. Each value of the time index corresponds to one of a number of time buckets: recent, adjacent periods of time of an equal length, such as previous hours, previous days, or previous weeks. The browsing history information may have a single browsing activity value for each time bucket, or it may have several depending upon other index dimensions of the array. The array may have an activity-type dimension, enabling the array to contain separate browsing activity scores for each of a number of different browsing activity types. Alternatively, all of the different browsing activities may be combined into a single weighted

browsing activity score. The array may also have an additional dimension indexed by item identifier, enabling the array to contain browsing activity scores for each of a number of different items offered for sale by the merchant.

[0022] In some cases, browsing activity scores each constitute a count of the number of times one or more browsing activities were performed within the time bucket for a particular item. In some cases, however, the score is a count of the number of unique users that performed the browsing action or actions. In some cases, the individual browsing action occurrences making up the score are weighted based upon the past history of the browsing action occurrences in successfully predicting orders for the item. Such weighting may be based on a variety of factors, including the type of browsing action, the identity of the user performing the browsing action, the identity of a referrer web site that sent the user to the merchant's web site, etc.

[0023] In transformation 202, the facility projects a browsing forecast 203 from the browsing history 201. This projection, as well as other projections discussed below, can be performed using a wide variety of statistical techniques, including projection techniques discussed in U.S. Patent Application No. 10/406,626, filed April 3, 2003, which is hereby incorporated by reference in its entirety. Projection techniques used by some embodiments of the facility include *moving average*, *exponential smoothing*, *Box-Jenkins ARIMA models*, two or more of which may be combined in hybrid approaches. The browsing forecast 203 is typically an array containing, for each of a number of future time buckets, a browsing activity score predicted for that future time bucket.

[0024] In transformation 204, the facility transforms the browsing forecast 203 into a purchasing forecast 205. The purchasing forecast is a projection of purchasing activity that will occur with respect to the item during each of a number of future time buckets. Such purchasing activity may take many forms, including adding the item to a shopping cart, checking out with the item in the shopping cart, initiating a one-click purchase for the item, providing payment information in connection with ordering the item, shipping the item, taking physical delivery of

the item by the purchaser, etc. A purchasing activity score may simply be a count of the expected number of occurrence of such purchasing actions, or may be a more complex weighted score based upon the numbers of such actions. In some embodiments, the conversion transformation is sensitive to such variables as time, item price, item availability, item sales cycle, and other sources of demand elasticity that affect the rate of conversion from browsing activity to purchasing activity. The facility projects a time-series of conversion ratios based on conversion history and some or all of the variables mentioned above. The facility applies this conversion ratio to the browsing forecast to produce a purchasing forecast.

[0025] This conversion projection may be generated as a function of the variables mentioned above, and not necessarily simply as numbers. As one example, output could be of the form  $\langle N, 0.3 \rangle$ ,  $\langle Y, 0.5 \rangle$  for each week in the forecast horizon where N indicates the item will not be available and Y indicates that it will be available. The above example considers only one variable, *availability*, but it can be extended to other variables and the representation of the conversion data could change accordingly. For instance, where price is a variable, embodiments of the facility use different values of conversion at different price tiers. Alternatively, the facility uses a function like:

$$c = 0.8 - (p - 10) * 0.1$$

Where the facility uses this function, for every dollar increase in price,  $p$ , above \$10, the conversion,  $c$ , drops by 0.1 points. In some embodiments, such functions are bounded to keep the conversion factors between 0 and 1.

[0026] In transformation 212, the facility uses purchasing history information 211 for the item to project a purchasing forecast 213. In transformation 221, the facility blends the purchasing forecast 205 from the browsing forecast and the purchasing forecast 213 from the purchasing history to obtain a blended purchasing forecast 222. The blending transformation determines, for each future time bucket, how heavily to weight the browsing activity score for that time bucket from each of the two purchasing forecasts in generating the blended purchasing

forecast. In some embodiments, in the blending transformation, the facility weights the purchasing forecast from browsing forecast more heavily where the item went out of stock at the merchant at least once during the past time buckets that make up the purchasing history, as the resulting unavailability of the item may have prevented customers that intended to purchase the item during those time buckets from doing so, thus influencing the purchasing activities measured in the purchasing history.

- [0027] In some embodiments, the facility performs blending using a function such as:

$$bl = a \cdot h + b \cdot f$$

where  $bl$  is the blended forecast,  $h$  is the purchasing forecast from the browsing forecast 205 and  $f$  is the purchasing forecast 213 from purchasing history for each time bucket.  $a$  and  $b$  are coefficients that typically satisfy the properties (1)  $0 \leq a, b$  and (2)  $a + b = 1$ , though in some situations, the second property is not maintained. Since  $h$  and  $f$  are inputs, the facility computes coefficients  $a$  and  $b$ . As in the case of the conversion model, these coefficients may also be generated as functions of the variables such as time, availability, price, etc. For instance, in some embodiments, the facility uses  $a=0$  and  $b=1$  if the item is expected to be out-of-stock in the future, and uses  $a=1$  and  $b=0$  if the item is expected to be available in the period under consideration. In some embodiments, the facility performs blending uniformly across all items, while in other embodiments the facility individually tailors the blending process somewhat to each item.

- [0028] Figure 3 is a graph showing the facility's generation of a sample blended purchasing forecast in accordance with the first approach. The X-axis of the graph shows time buckets, including feature time buckets having positive values and past time buckets having negative values. The origin on the X-axis shows what is roughly the present time. The Y-axis shows a quantitative indication of browsing and purchasing activity for the item, such as a raw number of browsing or purchasing actions, or a weighted score produced from the number of such actions performed.

- [0029] Time series 301 corresponds to browsing history 201, and is comprised of the following nine data points: (-9, 8), (-8, 10), (-7, 20), (-6, 27), (-5, 25), (-4, 29), (-3, 28), (-2, 30), and (-1, 27). For example, the first of these points indicates that, in the -9 time bucket (such as the 1-day time bucket occurring nine days before the present time), the browsing activity score for the item was 8, which might either correspond to a count of item browsing events, such as visits to the item's detail page, or may correspond to a weighted score generated from a wider variety of browsing actions. In accordance with projection transformation 202, time series 301 is transformed into time series 303, corresponding to browsing forecast 203. Time series 303 is in turn converted by conversion transformation 204 into time series 305, corresponding to purchasing forecast from browsing forecast 205.
- [0030] Time series 311 corresponds to purchasing history information 211. It can be seen that the purchasing activity scores in this time series are zero, both during past time buckets -9 through -8 and past time buckets -5 through -3. It may be that the item first became available for sale during bucket -7, making it impossible or unproductive to perform purchasing activities for the item in buckets -9 and -8. Additionally, the merchant's initial stock in the item may have been exhausted for time buckets -5 through -3, during which a purchasing activity score of zero was again registered. In the projection transformation 212, the facility transforms time series 311 into time series 313, corresponding to the purchasing forecast 213 from purchasing history. The facility then uses the blending transformation 221 to transform time series 305 and time series 313 into time series 322, corresponding to the blended purchasing forecast 222.
- [0031] The blended purchasing forecast represented by time series 322 may be used to anticipate future purchasing activity with respect to the item, and to set operating parameters with respect to the item such as inventory in the item, inventory in the item's complements, staffing levels among employees needed to sell the item, etc. Alternatively, the purchasing forecast represented by time

series 305, corresponding to purchasing forecast from processing forecast 205, may be used to anticipate future purchasing activity with respect to the item.

[0032] Figure 4 is a data flow diagram showing a second approach used by the facility to generate a blended purchasing forecast for an item. In the second approach, in projection transformation 402, the facility directly projects a purchasing forecast 405 from browsing history information 401. In a manner similar to the first approach, in the second approach, the facility projects a purchasing forecast 413 from purchasing history information 411. The purchasing forecast from browsing activity 405 and purchasing forecast from purchasing history 413 are combined in a blending transformation 421 to produce a blended purchasing forecast 422.

[0033] Figure 5 is a graph showing the facility's generation of a sample blended purchasing forecast in accordance with the second approach. In Figure 5, time series 501 corresponding to browsing history information 401 is transformed by projection transformation 402 into time series 505 corresponding to purchasing forecast information 405 from browsing history. Time series 511 corresponding to purchasing history 411 is transformed by projection transformation 412 into time series 513 corresponding to purchasing forecast information 413 from purchasing history. Time series 405 and time series 413 are combined by blending transformation 421 into time series 522, corresponding to blended purchasing forecast 422. The time series produced may be used in a manner similar to that described above in conjunction with the time series shown in Figure 3.

[0034] Figure 6 is a data flow diagram showing a third approach used by the facility to generate a purchasing forecast for an item. In the third approach, in projection transformation 631, the facility directly projects a unified purchasing forecast 632 from browsing history information 601 and purchasing history information 611. In some embodiments, purchasing history information 611 comprises a single time series containing the item's purchasing history, and the browsing history information 601 contains either (1) a single time series corresponding to all of the browsing history information associated with the item,

or (2) a number of different time series, each corresponding to the performance of a different browsing action performed with respect to the item. In some embodiments, projection transformation 631 uses a best-pick statistical technique, such as the versions of this technique described in Box et al., "Time Series Analysis: Forecasting & Control," Prentice Hall, 3<sup>rd</sup> Edition, February 28, 1994; Brockwell et al., "Introduction to Time Series and Forecasting," Springer Verlag, 2<sup>nd</sup> Book Edition, March 8, 2002; Hamilton, James D., "Time Series Analysis," Princeton University Press, January 11, 1994; Fuller, Wayne A., "Introduction to Statistical Time Series;" John Wiley & Sons, 2<sup>nd</sup> Edition, December 1995; and, Arsham, Hossein, "Time Series Analysis and Forecasting Techniques," February 18, 1994, <http://www.ubmail.ubalt.edu/~harsham/stat-data/opre330Forecast.htm>, each of which is hereby incorporated by reference in its entirety.

[0035] It will be appreciated by those skilled in the art that the above-described facility may be straightforwardly adapted or extended in various ways. For example, the facility may be used by a wide variety of merchants, and may project future purchasing activity of a variety of types based upon past browsing activity and/or purchasing activity that take a variety of forms and are observed in a variety of ways. While the foregoing description makes reference to preferred embodiments, the scope of the invention is defined solely by the claims that follow and the elements recited therein.